Speech Processing 2008/09

2nd Test

May 25th 2009

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)
   (a) In voiced sounds encoded with CELP, the contribution of the adaptive codeword is much smaller than the one of the stochastic codeword.
   (b) Subband coders are frequently used in audio coding.
   (c) Dual prediction residuals exhibit a non-flat spectral envelope.
   (d) The sinusoidal model matches sinusoids in consecutive frames using a nearest-neighbour frequency tracking method.
   (e) Segmental vocoders should be assessed in terms of speaker recognizability.
2. (a) What is the MOS achieved by the standard coding system adopted by the majority of fixed “plain old telephone services”?
   (b) How many bits per sample are necessary?
3. Give examples (one per class) of
   (a) a coder using vector quantization and low delay
   (b) a coder using QMF filters
   (c) a variable bit rate coder
   (d) a coder using mixed excitation
   (e) an analysis-by-synthesis coder with long-term prediction, without vector quantization
   (f) a coder which transmits voicing info per band
4. Let us consider a CELP type coder, operating with telephone bandwith (sampling frequency 8 kHz).
   (a) The short-term predictor computes 10 coefficients per frame, using windows of duration 22.5 ms, spaced 15 ms. If each coefficient is quantized with 3 bits, compute the bit rate (bps) assigned to this predictor.
   (b) The long-term predictor contributes with 1600 bps for the total bit rate. Specify a bit assignment (index and scale factor), and frame spacing (in ms) that justifies this bit rate.
   (c) The stochastic codeword is 7.5 ms long, and is quantized with 10 bits for the index and 5 for the scale factor. Compute its contribution (bps) for the total bit rate.
   (d) Compute the total bit rate (bps).
5. Classify as True (T) or False (F)
   (a) Mixed-excitation models are more adequate for representing breathy voices, than dual excitation models (pulse train or white noise).

(b) Epoch detection is the most sensitive part of prosody modifications in PSOLA methods.

(c) A concatenative synthesis system based on syllables has a smaller sound inventory than one based on diphones.

(d) Formant synthesizers are more flexible for generating voice effects than concatenative synthesizers.

(e) In formant synthesizers, target values for formant trajectories are not always reached depending on the phone duration and the phone context.

6. Indicate one strong point and one weak point about articulatory synthesizers.

7. How many break indices does the TOBI model define?

8. For which classes of sounds do formant synthesizers typically use the parallel model?

9. When synthesizing the word *café*, the prosody generation module computed a target duration of 120 ms for the fricative sound. The two diphones in the sound inventory that will be concatenated are: *6f* (with duration values 90 and 65 for the two sub-units) and *fE* (with duration values 85 and 70 for the two sub-units). Indicate the target durations for the two parts of the fricative, corresponding to the first and second diphones.

10. Write two pronunciations for the following foreign names. The first pronunciation should be the one typically heard in Portuguese media. The second pronunciation is the one obtained using the grapheme-to-phone rules for the common lexicon in European Portuguese. Both should use only the phonetic SAMPA symbols of this language.

    (a) Messenger

    (b) Google

    (c) Financial Times

11. Using the following syntax
    a → b / c .. d
    describe the simplified rules for grapheme *o* (unstressed) for Brazilian Portuguese, where c and d may be graphemes (e.g.: a, b, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (#), or any grapheme (*). You may use the symbol *0* to mark phonemic nulls, and the symbols | and ( ) to mark disjunction between several graphemes (e.g.: a | b | c). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The rules do not need to contemplate compound words. In fact, they should only account for the cases depicted in the examples below:
    Examples of transcriptions different from “o” (identical to European Portuguese): contar, comprar, ao, caos, farto, sentidos.
    Example of transcription “ow”: ouvir.
    Examples of transcription “o”: conhecer, correr, colorida.

Test 2 - Answers

Name:	
Number:	

1. (1.5 val.) Indicate T or F:

a	b	c	d	e

2. (1.2 val.)

a	
b	

3. (3.6 val.)

a	
b	
c	
d	
e	
f	

4. (1.5 val.)

a	
b	
c	
d	

5. (1.5 val.) Indicate T or F:

a	b	c	d	e

6. (1.6 val.)

strong point	
weak point	

7 to 9 (1.6/1.0/1.5 val.)

7	
8	
9	

10. (2 val.)

transc	Messenger	Google	Financial Times
transc1			
transc2			

The answer to the last question (11) should be written in the next page (3 val.).